

## University of Groningen

### Rational und Fair

Ockenfels, A.; Raub, W.

*Published in:*  
Kölner Zeitschrift für Soziologie und Sozialpsychologie

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2010

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
Ockenfels, A., & Raub, W. (2010). Rational und Fair. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 50, 119-136.

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## **RATIONAL UND FAIR<sup>†</sup>**

**Axel Ockenfels**  
Staatswissenschaftliches Seminar  
Universität zu Köln  
[ockenfels@uni-koeln.de](mailto:ockenfels@uni-koeln.de)

**Werner Raub**  
Institut für Soziologie/ICS  
Universität Utrecht  
[w.raub@uu.nl](mailto:w.raub@uu.nl)

Version 5. August 2009

Zeichen (inkl. Leerzeichen, Titelblatt und Literaturverzeichnis): 57494

Beitrag für:  
Gert Albert und Steffen Sigmund (Hrsg.)  
*Soziologische Theorie kontrovers*  
*Sonderheft 50(2010)*  
*Kölner Zeitschrift für Soziologie und Sozialpsychologie*

---

<sup>†</sup> Ockenfels dankt der Deutschen Forschungsgemeinschaft für finanzielle Unterstützung. Raub dankt der Niederländischen Organisation für wissenschaftliche Forschung (NWO) für finanzielle Unterstützung durch das PIONIER-Programm „The Management of Matches“ (S 96-168 und PGS 50-370) und das Projekt „Commitments and Reciprocity“ (400-05-089).

# RATIONAL UND FAIR

Axel Ockenfels und Werner Raub

*Zusammenfassung:* Das „Standardmodell“ des Rational Choices-Ansatzes, das Modell des *homo oeconomicus*, unterstellt rationales und eigennütziges Verhalten. Das Standardmodell hat manche Vorzüge, aber es gibt auch empirische Regelmäßigkeiten, die diesem Modell widersprechen. Wir behandeln experimentelle Evidenz zu Markt-Spielen, Diktator- und Ultimatumspielen und zum Gefangenendilemma. Wir skizzieren Alternativen zum Standardmodell, die an der Rationalitätsannahme festhalten, aber die Eigennutzannahme aufgeben bzw. modifizieren. Bei einer dieser Alternativen handelt es sich um ein Modell, in dem die Präferenzen der Akteure selbst Modellkonsequenzen und keine Modellannahmen sind. Ein zweites Modell unterstellt zusätzlich zum Eigennutz ein einfaches Motiv basierend auf sozialen Vergleichsprozessen, und kann so ein breites Spektrum von empirischen Regelmäßigkeiten organisieren, einschließlich solcher, die mit der Eigennutzannahme unvereinbar sind.

*Stichworte:* Assurance Game, Diktatorspiel, Dilemma-Spiele, Eigennutz, endogene Präferenzänderungen, Experimente, Fairness, homo oeconomicus, Markt-Spiele, Rationalität, Reziprozität, Spieltheorie, Ultimatumspiel

*Abstract:* The standard model of the rational choice approach, i.e., the *homo oeconomicus* model, assumes rational as well as self-interested behavior. This model has various strengths but also encounters empirical regularities that contradict various of its predictions. With respect to empirical evidence, we focus on experimental findings on market games, Ultimatum and Dictator Games, and the Prisoner's Dilemma. We sketch alternatives for the standard model that retain the rationality assumption but replace or modify the assumption of self-interest. These alternatives include a model of endogenous preference change with non-selfish preferences as *implications* rather than *assumptions* of the model. Another model includes the motivational assumption that actors care for their own material payoff and for their relative payoff standing and can account for a large and seemingly disparate set of empirical regularities, including regularities that contradict the self-interest assumptions.

*Keywords:* Assurance Game, Dictator Game, endogenous preference change, experiments, fairness, game theory, homo oeconomicus, market games, Prisoner's Dilemma, rationality, reciprocity, self-interest, ultimatum game

## 1. Das „Standardmodell“ rationalen und eigennützigen Verhaltens und seine Grenzen am Beispiel experimenteller Befunde

Man kann über beinahe alles streiten und in den Sozialwissenschaften wird auch über viel gestritten, aber es ist dem Erkenntnisfortschritt nicht förderlich, über alles zugleich zu streiten. Daher nehmen wir in diesem Beitrag an, dass es in den Sozialwissenschaften jedenfalls auch und wesentlich um die Erklärung sozialer Phänomene und Prozesse und um die Erklärung sozialer Regelmäßigkeiten geht. Erklärungen sind im Prinzip und idealerweise deduktive Argumente oder Varianten solcher Argumente. Wir nehmen weiter an, dass es in den Sozialwissenschaften jedenfalls auch und wesentlich um die empirische Überprüfung solcher Erklärungen geht: auch in den Sozialwissenschaften streben wir nach empirisch gehaltvollen und bewährten Erklärungen. Schließlich nehmen wir an, dass sozialwissenschaftliche Erklärungen elementaren heuristischen Regeln des methodologischen Individualismus folgen. Es geht also darum, soziale Phänomene und Prozesse – Makro-Phänomene wie z.B. Preise und Eigenschaften von Güterverteilungen wie deren „Fairness“ oder „Effizienz“ – als Resultate des Zusammenspiels individueller Handlungen zu erklären, wobei individuelle Handlungen – Mikro-Phänomene – ihrerseits durch soziale Bedingungen beeinflusst werden.<sup>1</sup> Unser Beitrag folgt damit einem sozialwissenschaftlichen Erkenntnisprogramm, wie es in inzwischen bereits klassischer Weise durch Hans Albert (vgl. z.B. 1977) beschrieben wurde. Albert hat vor allem die ökonomische Wissenschaft im Blick. Lindenberg und Wippler (z.B. 1978), Opp (z.B. 1979) und Esser (z.B. 1993) haben dieses Programm, das auf die schottische Moralphilosophie (David Hume, Adam Smith) zurück geht, auch den Soziologen wieder in Erinnerung gebracht. Die Wirkung der Arbeiten von Coleman (1990) und Hedström (2005) zeigt, dass dieses Erkenntnisprogramm in der modernen Soziologie inzwischen Einfluss gewonnen hat. Das nährt die Hoffnung, dass Ökonomen und Soziologen durch ein gemeinsames Erkenntnisprogramm fruchtbar zusammenarbeiten und voneinander lernen können.

Der Rational Choice-Ansatz ist eine Variante des skizzierten Erkenntnisprogramms. Das „Standardmodell“ dieses Ansatzes, gelegentlich auch als das Modell des *homo oeconomicus* angedeutet, unterstellt rationales und eigennütziges Verhalten. In unserem Beitrag beschäftigen wir uns mit Situationen strategischer Interdependenz zwischen zwei oder mehr Akteuren. Damit ist gemeint, dass die Entscheidungen eines Akteurs Folgen haben für den oder die „Mitspieler“ und umgekehrt. Nimmt man weiter an, dass die Akteure bei ihren Handlungen diese Interdependenzen auch „in Rechnung stellen“, dann sind wir bei Webers (1976[1921]: 1) berühmter Definition der Soziologie angelangt: „Soziologie ... soll heißen: eine Wissenschaft, welche soziales Handeln deutend verstehen und dadurch in seinem Ablauf und seinen Wirkungen ursächlich erklären will ... ‚Soziales‘ Handeln ... soll ein solches Handeln heißen, welches seinem von dem oder den Handelnden gemeinten Sinn nach auf das Verhalten anderer bezogen wird und daran in seinem Ablauf orientiert ist.“ Die Spieltheorie ist der Zweig der Theorie rationalen Handelns, der Situationen mit Interdependenzen zwischen Akteuren modelliert, und zwar unter Webers Annahme, dass die Akteure in der Tat ihre wechselseitige Beziehung und Abhängigkeit insofern in Rechnung stellen, dass sie auch das Handeln der anderen Akteure berücksichtigen. Die Spieltheorie ist damit ein maßgeschneidertes Instrument

---

<sup>1</sup> Es ist klar, dass unsere Beispiele für Makro-Phänomene typische Forschungsfelder von Ökonomen und Soziologen betreffen. In der Soziologie werden sie z.B. unter Stichworten wie „soziale Ungleichheit“ und „Problem der sozialen Ordnung“ behandelt (vgl. z.B. Diekmann und Voss 2004).

für eine Soziologie in Webers Sinn (siehe dazu Raub und Buskens 2006). In Situationen strategischer Interdependenz liegt es nahe, rationales Verhalten als Gleichgewichtsverhalten zu interpretieren: jeder Akteur maximiert seinen erwarteten Nutzen, gegeben die Strategien aller anderen Akteure.<sup>2</sup>

Zufolge der Rationalitätsannahme maximieren Akteure ihre Nutzenfunktion, aber die Rationalitätsannahme als solche sagt nichts darüber, was Akteure „inhaltlich“ nützlich finden. Darüber wird man aber häufig Annahmen machen müssen, will man Verhalten erklären. Das Standardmodell des Rational Choice-Ansatzes unterstellt daher nicht nur rationales Verhalten, sondern es enthält auch eine Annahme über das Argument der Nutzenfunktion, also über die Motivation der Akteure, nämlich die Annahme eigennützigem Verhalten: Akteure sind ausschließlich durch den eigenen materiellen Gewinn motiviert („Nutzen = eigenes Geld“) und mehr Geld ist besser als weniger Geld.

Um Verhalten erklären und prognostizieren zu können, müssen neben den individuellen Motivationen häufig institutionelle und soziale Rahmenbedingungen spezifiziert werden, etwa der Informationsstand der Akteure und andere Umstände, unter denen die Interaktion stattfindet. Es ist nicht ganz unüblich, gerade in manchen „soziologienahen“ Darstellungen, im Zusammenhang mit sehr spezifischen Annahmen, z.B. vollständige Informiertheit und perfekte, anonyme Märkte, auf die neoklassische Mikro-Ökonomik als eine Disziplin zu verweisen, in der solche Annahmen typischerweise verwendet werden. Dabei sollte man sich allerdings darüber im Klaren sein, dass es inzwischen schon recht lange her ist, dass sich die neoklassische Mikro-Ökonomik in ihrer Breite einigermaßen treffend auf diese Weise charakterisieren ließ (man konsultiere etwa eine maßgebende Einführung wie Mascoll et al. 1995, um zu sehen, dass die moderne Mikro-Ökonomik sehr viel mehr ist als allgemeine Gleichgewichtstheorie für perfekte Märkte).

Das Standardmodell mit seinen Annahmen Rationalität und Eigennutz hat wichtige Vorzüge. Zu diesen Vorzügen gehört, dass es ein einfaches und sparsames Modell mit hohem Informationsgehalt und hoher empirischer Überprüfbarkeit im Sinn Poppers (1973[1934]) ist: die Kombination von Rationalitäts- und Eigennutzannahme führt zu vielen empirisch prüf- und widerlegbaren Implikationen. Diese Vorzüge entgehen auch denjenigen Kritikern des Standardmodells nicht, die wissenschaftstheoretisch geschult sind und ansonsten das oben skizzierte sozialwissenschaftliche Erkenntnisprogramm teilen (vgl. Abschnitt 2 von Essers Beitrag in diesem Band für ein gutes Beispiel). Es gibt auch andere Argumente für das Standardmodell. Eines dieser Argumente ist interessant, weil es darauf abhebt, dass es in der Soziologie und in anderen sozialwissenschaftlichen Disziplinen anders als in der Psychologie ja nicht um die Erklärung individuellen Verhaltens an sich geht, sondern um die Erklärung von Makro-Phänomenen, die aus individuellem Verhalten resultieren. Ganz im Sinn dieses Gedankens finden wir in der Ökonomik bei Becker (1976: Kap. 9) und in der Soziologie bei Coleman (1987) und Goldthorpe (1996) das Argument, dass man auf der Mikro-Ebene individuellen Verhaltens zwar möglicherweise Widerlegungen des Standardmodells antreffe („Anomalien“) und dass komplexere Theorien dem Standardmodell möglicherweise überlegen seien im Hinblick auf die Erklärung individuellen Verhaltens, dass solche komplexeren Theorien aber gerade wegen ihrer Komplexität ungeeignet seien für die Ableitung von Makro-Implikationen oder aber, dass die Makro-Implikationen komplexerer

---

<sup>2</sup> Wir verzichten in diesem Beitrag auf die Diskussion technischer Details. Vgl. Diekmann (2009) für eine moderne und leicht zugängliche Einführung in die Spieltheorie (dort auch Hinweise auf weiterführende und technisch anspruchsvollere Literatur).

Individualtheorien sich jedenfalls nicht von denen des Standardmodells unterscheiden.<sup>3</sup>

In diesem Beitrag konzentrieren wir uns auf experimentelle empirische Evidenz zum Standardmodell. Uns geht es hier um empirische Evidenz für oder gegen alternative Theorien und Hypothesen. Die Vorzüge experimenteller Designs bei der Überprüfung von Theorien und Hypothesen – im Gegensatz zur Suche nach Zusammenhängen und empirischen Regelmäßigkeiten – sind bekannt, gerade dann, wenn zusätzlich im Rahmen von Kreuzvalidierungen Laborexperimente mit Feldexperimenten und auch mit nicht-experimentellen Methoden wie etwa Surveystudien verbunden werden (vgl. z.B. Ockenfels 2009 sowie Buskens und Raub 2010), um so den verschiedenen Problemen experimenteller Designs zu begegnen. Bei der Überprüfung und Weiterentwicklung von Theorien im Rahmen des Rational Choice-Ansatzes spielen Experimente inzwischen eine immer größere Rolle (vgl. z.B. Camerer 2003 über die Forschungsrichtung „Verhaltensökonomik“). Es gibt Anzeichen, dass die systematische Verwendung der experimentellen Methode eine wichtige Rolle spielen kann bei der Annäherung von Ökonomik und Soziologie im Rahmen eines gemeinsamen Erkenntnisprogramms (vgl. z.B. Diekmann 2008 und Fehr und Gintis 2007).

Das Standardmodell ist oft in der Lage, das Verhalten von vielen Individuen mit unterschiedlichen Interessen und Ausgangspositionen selbst in hochkomplexen Situationen adäquat abzubilden (und ist daher beispielsweise in der Wettbewerbspolitik sowie im angewandten Marktdesign ein in der Praxis zunehmend akzeptierter Ansatz). Ein einfaches Beispiel ist das *Marktspiel* von Roth et al. (1991), in dem 9 Käufer simultan einem Verkäufer jeweils ein Gebot für ein unteilbares Gut machen. Das Gut besitzt für alle Käufer denselben Wert (\$10), aber keinen Wert für den Verkäufer. Der Verkäufer kann das höchste Gebot annehmen oder ablehnen. Im letzteren Fall erhält keiner der Akteure eine positive Auszahlung. Wenn der Verkäufer akzeptiert, erhält er das höchste Gebot, und der Käufer mit dem höchsten Gebot erhält die Differenz zwischen dem Wert und seinem Gebot. Wenn mehrere Käufer den höchsten Preis bieten, entscheidet das Los. Alle anderen Käufer gehen leer aus. Ist mehr Geld besser als weniger, nimmt der Verkäufer jedes positive Gebot an. Jeder Käufer, der nicht das höchste Gebot macht, erhält mit Sicherheit \$0, während die Käufer mit dem höchsten Gebot eine positive erwartete Auszahlung haben, wenn das Gebot kleiner ist als \$10. Folglich gilt, dass alle Käufer dasselbe Gebot machen, wenn das höchste Gebot kleiner als \$10 ist. Wenn aber alle Käufer ein Gebot kleiner als \$10 machen, so ist es für einen Käufer vorteilhaft, sein Gebot ein wenig anzuheben, so dass er den Zuschlag mit Sicherheit anstatt lediglich mit Wahrscheinlichkeit 1/9 erhält. Folglich muss in jedem Gleichgewicht ein Angebot \$10 sein und vom Verkäufer auch angenommen werden, so dass der Verkäufer den gesamten Wert des Gutes abschöpft, während alle Käufer leer ausgehen. Tatsächlich zeigt sich, dass diese

---

<sup>3</sup> Eine gewisse Verwandtschaft dieser Argumente mit Friedmans (1953) „as if“-Interpretation des Standardmodells ist offensichtlich. Und natürlich gibt es auch noch andere Argumente für das Standardmodell. Dazu gehört der Vorschlag, das Standardmodell als „worst case“-Szenario für institutionelles Design und wirtschaftspolitische Entscheidungen zu verwenden (vgl. Brennan und Buchanan 1985: Kap. 4 für eine systematische Ausarbeitung dieses Gedankens und Schüßler 1988 für eine kritische Diskussion). Dazu gehört auch die Verwendung des Standardmodells als „benchmark“, um danach empirisch beobachtete Abweichungen von den Vorhersagen des Standardmodells als Explananda zu behandeln (vgl. z.B. Harsanyi 1976: Kap. 6, 7; 1977: Kap. 1, 2). Eine nähere Diskussion dieser Argumente würde aber den Rahmen unseres Beitrags sprengen. Viele Vertreter des Rational Choice-Ansatzes anerkennen nicht nur dessen Grenzen als Erklärungsansatz, sie sind auch selbst maßgeblich daran beteiligt, diese auszuloten; wir kommen darauf noch zurück.

spieltheoretische Prognose eine hohe Anziehungskraft besitzt: in allen Märkten in allen vier Ländern, in denen Roth et al. (1991) das Experiment durchgeführt haben, wird das Gleichgewicht bereits nach wenigen Runden erreicht, obwohl es zu extrem ungleichen Auszahlungsverteilungen führt. Dieses und ähnliche Marktspiele wurden von zahlreichen Forschern repliziert, mit stets denselben Ergebnissen und Schlussfolgerungen.

Dieselbe Studie von Roth et al. (1991) untersucht auch Ultimatumspiele, die erstmals von Güth et al. (1982) experimentell untersucht wurden. Das Ultimatumspiel ist den formalen Regeln nach ein Marktspiel mit nur einem Käufer, der hier „Proposer“ genannt wird. Der Proposer schlägt also einem (anonymen) „Responder“ (dem Verkäufer im Marktspiel) die Aufteilung eines Geldbetrages vor. Der Responder hat dabei ein Vetorecht, d.h. er kann das Angebot annehmen oder ablehnen. Nimmt er es an, erhalten beide Akteure die Auszahlung, die der Proposer vorgeschlagen hat. Lehnt der Responder dagegen ab, so erhält keiner der beiden Akteure eine positive Auszahlung. Unter der Annahme, dass beide Akteure mehr Geld weniger Geld vorziehen, sollte der Responder jedes positive Angebot annehmen. Im (teilspielperfekten) Gleichgewicht wird der Proposer konsequenterweise nicht mehr als die kleinste mögliche Geldeinheit anbieten. Wie im Marktspiel kommt es also auch in diesem elementaren Verhandlungsspiel im Gleichgewicht zu einer extrem ungleichen Auszahlungsverteilung. Doch anders als im Marktspiel versagt das Standardmodell. Im Ultimatumspiel werden nur selten geringe Angebote gemacht, und wenn doch, werden diese oft abgelehnt. Der Modalwert für das Angebot liegt bei einer Gleichaufteilung der Auszahlungen. Auch dieses Ergebnis hat sich in hunderten Variationen als außerordentlich robust erwiesen (siehe Roth 1995 für einen Übersichtsartikel über die frühe Literatur).

Das Diktatorspiel besitzt dieselben Regeln wie das Ultimatumspiel mit der Ausnahme, dass der Responder kein Vetorecht besitzt: der Proposer bestimmt schlicht die Aufteilung eines Geldbetrags. Ist mehr Geld besser als weniger Geld, sollte der Proposer alles für sich behalten. Doch auch hier geben viele etwas ab, und einige wählen gar die Gleichaufteilung – allerdings signifikant weniger als im Ultimatumspiel (Forsythe et al. 1994).

Ähnlich robuste Abweichungen von eigennützigen Strategien werden auch im Gefangendilemma beobachtet, in dem keiner von zwei eigennützigen Akteuren kooperieren sollte, egal was der Partner macht, und dies obwohl wechselseitige Kooperation beiden Akteuren eine höhere (und faire) Auszahlung bringen würde. Typischerweise beobachtet man jedoch, dass der Anteil an kooperativen Entscheidungen signifikant über Null liegt, dass Kooperation auf die Kooperationsbereitschaft des Partners konditioniert wird (Reziprozität), dass aber effiziente Ergebnisse oft ebenso deutlich unerreicht bleiben (vgl. auch Kapitel 3 sowie die Übersichtsartikel Roth 1988, und Dawes und Thaler 1988).

Insgesamt zeigen sich in den verschiedenen Kontexten – Märkte, Verhandlungen, soziale Dilemmata – sehr unterschiedliche Verhaltensweisen: wettbewerbliches, eigennütziges Verhalten in Märkten, Fairness bei Verhandlungen und Reziprozität bei Kooperationsproblemen. Das Standardmodell liefert außerhalb von Märkten oft unbefriedigende Erklärungen für Mikro- und Makro-Phänomene.

## 2. Alternativen zum Standardmodell: Aufgabe der Eigennutzannahme

Die skizzierte experimentelle Evidenz motiviert die Entwicklung von theoretischen Alternativen zum Standardmodell. Man kann dazu unterschiedliche Wege einschlagen (vgl. z.B. Ockenfels 1999: Kap. 2 für eine knappe Übersicht zu alternativen Modellen für die Erklärung empirischer Phänomene in den drei Kontexten, die wir im vorhergehenden Abschnitt besprochen haben). Man kann insbesondere unterscheiden zwischen der Aufgabe oder Modifikation der Rationalitätsannahme, der Aufgabe oder Modifikation der Eigennutzannahme und der Aufgabe oder Modifikation der Annahmen über die sozialen Kontexte, unter denen sie handeln. Es empfiehlt sich dabei, systematisch und schrittweise vorzugehen und nicht alle Annahmen des Standardmodells zugleich fallen zu lassen. Durch schrittweises Vorgehen kann man hoffen, mehr darüber zu erfahren, *welche* Annahmen des Standardmodells unter *welchen* Bedingungen mehr oder weniger problematisch sind.

Manche empirische Evidenzen der skizzierten Art lassen sich dadurch erklären, dass man anstelle der Rationalitätsannahme z.B. Lernmodelle verwendet (vgl. z.B. die Übersicht in Camerer 2003: Kap. 6). Andere empirische Evidenzen werden verständlich, wenn man soziale Bedingungen wie z.B. wiederholte Interaktionen und Netzwerke von Beziehungen zwischen Akteuren berücksichtigt und gegebenenfalls zusätzlich in bestimmten Hinsichten unvollständige Informationen der Akteure über ihre Interaktionspartner annimmt. Unter solchen Bedingungen wird ansonsten rationales und eigennütziges Verhalten auch dadurch beeinflusst, dass die Akteure die langfristigen Folgen ihres Handelns (Reziprozität im Sinne bedingter Kooperation als Resultat „aufgeklärten Eigeninteresses“) und Reputationseffekte berücksichtigen (vgl. als Übersicht z.B. Buskens und Raub 2010 und Bolton und Ockenfels, im Erscheinen a). Wir beschäftigen uns hier ausschließlich mit Modellen, die die Eigennutzannahme aufgeben, an der Rationalitätsannahme hingegen festhalten. Ein wesentlicher Grund dafür ist, dass das Standardmodell gerade auch in solchen Kontexten widerlegt wird, in denen es nicht nur für den Forscher, sondern auch für die Akteure selbst offensichtlich ist, welches Verhalten aus der Annahme rationalen und eigennützigen Verhaltens folgt, wobei es sich außerdem um Kontexte handelt, in denen die Akteure einmalig interagieren, in denen Netzwerkeffekte keine Rolle spielen und in denen auch unvollständige Information über den oder die Interaktionspartner das Verhalten eines rationalen und eigennützigen Akteurs nicht beeinflussen kann: so ist Defektion im nicht-wiederholten Gefangenendilemma für einen Akteur bei jeder Entscheidung des anderen Akteurs stets mit einer höheren eigenen Auszahlung verbunden. Die Konsequenzen für die eigene Auszahlung im Ultimatumspiel bei Ablehnung eines positiven Angebots des Proposers sind für den Responder klar und das gleiche gilt für den Diktator bei einer positiven Abgabe im Diktatorspiel. Es ist wenig plausibel, dass eigennützige Akteure in solchen Situationen allererst lernen müssen, sich rational zu verhalten und es kann auch nicht so sein, dass rationale und eigennützige Akteure in derartigen Kontexten durch die Antizipation von langfristigen Folgen und Reputationseffekten ihres Verhaltens beeinflusst werden. Die Akteure *wollen* vom homo oeconomicus Modell abweichen.

Wenn man die Eigennutzannahme des Standardmodells aufgibt und mithin andere und zusätzliche „Nutzenargumente“ berücksichtigt, dann ist das aus wissenschaftstheoretischer Sicht nicht unproblematisch: die Theorie wird komplexer und im Prinzip schwieriger widerlegbar. Wenn man die Motivationsannahmen „richtig“ oder komplex genug wählt, dann kann man zwar dafür sorgen, dass mehr oder weniger jedes Verhalten mit der Rationalitätsannahme und entsprechenden



Motivationsannahmen konsistent ist, aber gerade deshalb sind solche „Erklärungen“ aus wissenschaftstheoretischer Sicht wertlos (vgl. zu Vorzügen „einfacher“ Theorien allgemein Popper 1973[1934]: Kap. VI und neuer Anhang VIII; im Hinblick auf die Motivationsannahmen im Rahmen des Rational Choice-Ansatzes hat z.B. Harsanyi 1976: Kap. VII und vor allem S. 122 den Punkt klar formuliert; vgl. auch Esser in diesem Band und im Hinblick auf spieltheoretische Modelle Diekmann 2008: 543, sowie Schmidt 2009).

Man kann dem Problem von ad hoc-Erklärungen und „Immunisierungen“ (Albert 1977) des Rational Choice-Ansatzes durch Verwendung anderer Motivationsannahmen als der Eigennutzannahme auf mindestens zweierlei Weise entgehen. Eine Möglichkeit besteht darin, Alternativen zur Eigennutzannahme überhaupt nicht als Annahmen („Axiome“) zu verwenden, sondern in einem ersten Schritt ihrerseits als Konsequenzen aus einem allgemeineren Modell abzuleiten und mithin, wenn man so will, Bedingungen explizit zu spezifizieren, unter denen Akteure nicht oder jedenfalls nicht nur durch Eigennutz motiviert sind. Eine andere Möglichkeit ist, auch *neue* und *falsifizierbare* Vorhersagen aus alternativen Motivationsannahmen abzuleiten oder jedenfalls zu zeigen, dass man mit „sparsamen“ alternativen Motivationsannahmen sehr verschiedene und prima facie „widersprüchliche“ empirische Evidenzen erklären kann, die mit den Vorhersagen des Standardmodells z.T. übereinstimmen, die anderen Vorhersagen des Standardmodells aber widersprechen. Im Hinblick auf die oben skizzierten experimentellen Befunde hieße dies z.B. zu zeigen, dass relativ einfache Alternativen zur Eigennutzannahme ausreichen, um sowohl augenscheinlich eigennütziges Verhalten auf Märkten als auch faires und reziprokes Verhalten in Fairness- und Dilemma-Spielen zu erklären (ganz im Sinn von Poppers 1963: 240-242 drei Kriterien für den Erkenntnisfortschritt: Einfachheit, unabhängige Prüfbarkeit und empirische Bewährung).

Im Folgenden skizzieren wir zwei theoretische Modelle. Eines dieser beiden Modelle ist eine Implementation des Gedankens, Alternativen zur Eigennutzannahme ihrerseits aus einem allgemeineren Modell abzuleiten. Das andere Modell folgt der zweiten Strategie und erklärt eine breite Palette von Verhaltensweisen in unterschiedlichen Kontexten mit relativ einfachen, aber von der Eigennutzannahme abweichenden Motivationsannahmen. Für beide Modelle gibt es in der Literatur Konkurrenten, die auf ähnlichen Überlegungen beruhen. Es geht uns in diesem Beitrag nicht um Belege, dass die hier skizzierten Modelle ihren Konkurrenten überlegen sind. Es geht uns lediglich um den Nachweis, dass die Aufgabe der Eigennutzannahme nicht notwendigerweise mit aus methodologischer Sicht problematischen ad hoc-Modifikationen des Rational Choice-Ansatzes verbunden sein muss.

### **3. Alternativen zur Eigennutzannahme als deduktive Konsequenzen eines Rational Choice-Modells**

Der Gedanke, Präferenzen zu „endogenisieren“, also Aussagen über Motivationen von Akteuren nicht als *Annahmen* eines Modells zu verwenden, sondern solche Aussagen als *Konsequenzen* eines Modells abzuleiten, ist im Rahmen des Rational Choice-Ansatzes häufiger anzutreffen. Dabei begegnen wir auch verschiedenen Ansätzen, in denen die Motivationen selbst Resultate rationalen Verhaltens sind, in denen Akteure also ihre Motivationen selbst in irgendeiner Weise wählen (vgl. z.B. Raub und Voss 1990 für eine Literaturübersicht). Wir skizzieren ein stark vereinfachtes Beispiel eines

solchen Ansatzes für den Spezialfall des Gefangenendilemmas, also für das Spiel in Abbildung 1.

Abbildung 1 ungefähr hier.

Wir wollen annehmen, dass die Parameter  $T$ ,  $R$ ,  $P$  und  $S$  die materiellen Auszahlungen für die beiden Akteure repräsentieren. Wenn also etwa Akteur 1 defektiert und Akteur 2 kooperiert, dann erhält Akteur 1 einen Geldbetrag  $T$  und Akteur 2 einen Geldbetrag  $S$ . Man nehme nun weiter an, dass beide Akteure rational und eigennützig sind. Dann gilt für ihre kardinalen Nutzenfunktionen  $u(T) = T^n > u(R) = R^n > u(P) = P^n > u(S) = S^n$  und sie würden beide defektieren, da Defektion dominante Strategie ist.<sup>4</sup> Nun nehmen wir an, dass die Akteure ihre eigenen Präferenzen wählen und gegebenenfalls auch modifizieren können. Wir modellieren die Wahl von Präferenzen als ein Spiel  $\mathbf{M}$  zwischen den beiden Akteuren. In diesem Spiel  $\mathbf{M}$  wählt jeder Akteur seine eigene effektive Präferenzordnung  $O_i$  ( $i = 1, 2$ ) für die materiellen Auszahlungen im Gefangenendilemma aus Abbildung 1. Wir nehmen an, dass  $\mathbf{M}$  ein nichtkooperatives Spiel ist. Die beiden Akteure können also keine bindenden Vereinbarungen über die Wahl ihrer Präferenzen treffen. Wir nehmen weiter an, dass die beiden Akteure in  $\mathbf{M}$  ihre effektiven Präferenzen simultan wählen. Effektive Präferenzen repräsentieren wir durch kardinale Nutzenfunktionen. Eine Strategie  $O_i$  von Akteur  $i$  in  $\mathbf{M}$  ist dann eine Ordnung der kardinalen effektiven Nutzenwerte  $T_i^e$ ,  $R_i^e$ ,  $P_i^e$  und  $S_i^e$  für die materiellen Auszahlungen  $T$ ,  $R$ ,  $P$  und  $S$  im 2x2-Spiel in Abbildung 2. Uns interessiert, ob rationale und eigennützige Akteure, die in der Lage sind, ihre eigenen Präferenzen zu modifizieren, gegebenenfalls andere als eigennützige Präferenzen wählen würden.

Abbildung 2 ungefähr hier.

Der Einfachheit halber wollen wir annehmen, dass jeder der beiden Akteure in  $\mathbf{M}$  zwischen zwei effektiven Präferenzordnungen bzw. Strategien wählen kann, die wir mit  $O_i^D$  und  $O_i^{AG}$  notieren. Die Präferenzordnung  $O_i^D$  entspricht eigennützigen Präferenzen, mithin gilt  $T_i^e = T^n > R_i^e = R^n > P_i^e = P^n > S_i^e = S^n$ . Die andere Präferenzordnung  $O_i^{AG}$  entspricht dagegen Sens (1982: Kap. 3) Assurance Game Präferenzen und es gilt  $R_i^e > T_i^e > P_i^e > S_i^e$ . Unter dieser Präferenzordnung ist Defektion im effektiven Spiel nicht mehr dominante Strategie. Vielmehr bleibt eigene Defektion zwar (einzige) beste Antwort gegen Defektion des anderen Akteurs, aber eigene Kooperation ist nun (einzige) beste Antwort gegen Kooperation des anderen Akteurs. Haben beide Akteure Assurance Game Präferenzen im effektiven Spiel in Abbildung 2, dann ist also nicht nur wechselseitige Defektion ein Gleichgewicht, sondern auch wechselseitige Kooperation. Wechselseitige Kooperation stellt nicht nur jeden der beiden Akteure besser als wechselseitige Defektion, sondern ist sogar für jeden der beiden Akteure mit dem höchsten möglichen Nutzen verbunden.

Nun müssen wir die Nutzenfunktionen der Akteure für das Spiel  $\mathbf{M}$  spezifizieren. Dazu machen wir uns klar, dass eine Strategienkombination  $O = (O_1, O_2)$  im Spiel  $\mathbf{M}$  die Nutzenfunktionen der Akteure für das effektive Spiel in Abbildung 2 festlegt. Wir nehmen an, dass das effektive Spiel in Abbildung 2 ein

<sup>4</sup> Um den Vergleich zu erleichtern, verwenden wir Notation aus Raub und Voss (1990). Dort deutet das Superskript „n“ die „natürlichen Neigungen“ der Akteure an. In unserem Modell nehmen wir an, dass diese natürlichen Neigungen eigennütziger Art sind. Die Pointe des Modells ist dann gerade der Nachweis, dass Akteure auch unter der (möglicherweise „pessimistischen“) Annahme, dass sie eigennützig sind, Anreize haben, ihre eigennützigen Präferenzen zu modifizieren.

nichtkooperatives Spiel mit vollständiger Information ist. Das bedeutet insbesondere, dass jeder der beiden Akteure nicht nur über seine eigenen effektiven Präferenzen informiert ist, sondern auch über die effektiven Präferenzen des anderen Akteurs, nachdem diese effektiven Präferenzen gewählt wurden und bevor das effektive Spiel selbst gespielt wird. Die zentrale Idee für die Spezifikation der Nutzenfunktionen für das Spiel **M** ist nun, dass beide Akteure auf die Folgen der in **M** gewählten effektiven Präferenzen für das Verhalten im effektiven Spiel aus Abbildung 2 antizipieren, wobei sie diese Folgen *eigennützig* bewerten.

Nehmen wir also an, dass beide Akteure im Spiel **M** die Präferenzordnung  $O_i^D$  wählen, also ihre eigennützigen Präferenzen nicht modifizieren. Das effektive Spiel in Abbildung 2 hat dann nicht nur im Hinblick auf die Ordnung der materiellen Auszahlungen die Struktur eines Gefangenendilemmas, es ist auch ein Gefangenendilemma im Hinblick auf die Präferenzen der Akteure und sie werden unter der Rationalitätsannahme im effektiven Spiel beide defektieren. Wenn also beide Akteure im Spiel **M** die Präferenzordnung  $O_i^D$  wählen, ihre eigennützigen Präferenzen also nicht modifizieren, und wenn sie die Folgen der Wahl von Präferenzen im Spiel **M** für die Resultate des effektiven Spiels eigennützig bewerten, dann ist beiderseitige Wahl der Präferenzordnung  $O_i^D$  für beide Akteure mit dem Nutzenwert  $P^n$  im Spiel **M** verbunden.

Nun betrachten wir den Fall, dass beide Akteure im Spiel **M** die Präferenzordnung  $O_i^{AG}$  wählen, so dass  $R_i^e > T_i^e > P_i^e > S_i^e$  für das effektive Spiel gilt. Das effektive Spiel ist dann also selbst ein Assurance Game. Unter der relativ schwachen Annahme, dass rationale Akteure ein Gleichgewicht spielen, dass jeden Akteur besser stellt als alle anderen möglichen Strategienkombinationen und insbesondere auch alle anderen Gleichgewichte, werden die beiden Akteure in diesem effektiven Spiel also wechselseitig kooperieren. Mithin ist beiderseitige Wahl der Präferenzordnung  $O_i^{AG}$  für beide Akteure mit dem Nutzenwert  $R^n$  im Spiel **M** verbunden.

Es bleibt der Fall, dass einer der beiden Akteure im Spiel **M** die Präferenzordnung  $O_i^D$  wählt, also seine eigennützigen Präferenzen nicht modifiziert, während der andere Akteur im Spiel **M** die Präferenzordnung  $O_j^{AG}$  wählt. Im effektiven Spiel ist dann Defektion dominante Strategie für Akteur  $i$ , während Defektion einzige beste Antwort von Akteur  $j$  gegen Defektion von Akteur  $i$  ist. Wechselseitige Defektion ist mithin das einzige Gleichgewicht dieses effektiven Spiels. Wenn also einer der beiden Akteure im Spiel **M** die Präferenzordnung  $O_i^D$  wählt, also seine eigennützigen Präferenzen nicht modifiziert, während der andere Akteur im Spiel **M** die Präferenzordnung  $O_j^{AG}$  wählt, dann realisieren beide Akteure den Nutzenwert  $P^n$  im Spiel **M**.

Abbildung 3 ungefähr hier.

Abbildung 3 zeigt die Normalform von Spiel **M**. In diesem Spiel ist wechselseitige Wahl von Assurance Game Präferenzen  $O_i^{AG}$  ebenso ein Gleichgewicht wie wechselseitige Wahl von eigennützigen Präferenzen  $O_i^D$ . Wechselseitige Wahl von Assurance Game Präferenzen stellt jeden Akteur besser als alle anderen möglichen Strategienkombinationen und insbesondere auch besser als wechselseitige Wahl von eigennützigen Präferenzen. Die Wahl von Assurance Game Präferenzen ist sogar schwach dominante Strategie im Spiel **M**. Wir können daher annehmen, dass rationale Akteure im Spiel **M** Assurance Game Präferenzen wählen und im effektiven Spiel kooperieren. Wir sehen also, dass rationale und eigennützige Akteure, die ihre

Präferenzen selbst modifizieren können und die dabei von *eigennützigen* Motiven geleitet werden, in der Tat ihre Präferenzen anpassen.

Auf den ersten Blick liegt der Einwand nahe, dass wir hier einen extremen Spezialfall betrachtet haben. Wir haben z.B. angenommen, dass die beiden Akteure im Spiel **M** nur zwischen  $O_i^D$  und  $O_i^{AG}$  wählen können. Wenn man aber überhaupt annehmen will, dass Akteure ihre Präferenzen selbst wählen können, warum sollte man dann nicht auch noch ganz andere Präferenzordnungen als wählbar unterstellen? Gilt dann noch stets, dass Akteure ihre Präferenzen modifizieren und im effektiven Spiel kooperieren? Unser Spezialfall wäre wenig interessant, falls sich das Resultat nicht verallgemeinern lässt. Diese skeptische Reaktion ist berechtigt. Unser Resultat, nämlich die rationale (auf Gleichgewichtsverhalten beruhende) und eigennützige Modifikation von Präferenzen derart, dass die endogenen Präferenzordnungen Kooperation induzieren, lässt sich aber verallgemeinern. Wir können hier nicht auf technische Details eingehen (vgl. für ausführliche Analysen und technische Bedingungen und Einschränkungen Raub und Voss 1990 und Raub 1990), man kann aber grob gesprochen zeigen, dass sich das Resultat in ähnlicher Form auch dann ergibt, wenn im Spiel **M** beliebige effektive Präferenzordnungen gewählt werden können und dass das Resultat nicht nur für das Gefangenendilemma gilt, sondern auch für andere soziale Dilemmas, auch für solche mit mehr als zwei Akteuren.

In dem skizzierten Modell werden verschiedene restriktive Annahmen verwendet. Insbesondere haben wir angenommen, dass die Akteure vollständig informiert sind über die gewählten effektiven Präferenzen, also nicht nur ihre eigenen effektiven Präferenzen kennen, sondern auch die effektiven Präferenzen des anderen Akteurs, und dass sie sich im effektiven Spiel auch entsprechend ihren effektiven Präferenzen verhalten, dass sie sich also auf ihre effektiven Präferenzen „festlegen“ können (vgl. Raub und Voss 1990 für eine genauere Übersicht und vor allem auch für soziale Bedingungen, unter denen solche Annahmen einigermaßen plausibel erscheinen). Das ändert aber nichts an dem Nachweis, dass es im Prinzip möglich ist, Präferenzänderungen im Rahmen eines Rational Choice-Modells zu endogenisieren und Alternativen zur Eigennutzannahme als Konsequenzen eines Rational Choice-Modells zu betrachten.<sup>5</sup>

#### **4. ERC: neue Vorhersagen aus Alternativen zur Eigennutzannahme**

Das Ziel des ERC-Modells<sup>6</sup> ist, mit Hilfe einer empirisch fundierten Motivationsstruktur, Laborverhalten auf ein einfaches Prinzip zurückzuführen. Dabei greift ERC klassische Ideen der Soziologie und Sozialpsychologie auf (vgl. z.B. „Eminenz“ bei Hobbes, aber auch Bezugsgruppentheorie z.B. bei Merton), nämlich dass Akteure nicht nur durch die Resultate ihres Handelns für die eigene Person selbst beeinflusst werden, sondern auch durch „soziale Vergleichsprozesse“ (ein ähnliches Modell mit weitgehend analogen Resultaten wurde von Fehr und Schmidt 1999 entwickelt). Der Kern von ERC ist die Annahme, dass jeder Akteur *i* so handelt, als

---

<sup>5</sup> Ein alternativer Ansatz zur Endogenisierung sozialer Präferenzen ist der so genannte indirekt evolutionäre Ansatz, bei dem sich in einem formalen Modellrahmen die Präferenzen rationaler Individuen (sowie Aspekte der Institutionen) im evolutionären Wettbewerb anpassen können. Hier zeigt sich analog zur obigen Analyse, dass die erfolgreichen, im evolutionären Wettbewerb effektiv selektierten Präferenzen nicht etwa eigennütziger sondern reziproker Natur sind; siehe etwa Güth und Kliemt (2000) sowie Güth und Ockenfels (2003, 2005).

<sup>6</sup> E steht für Equity, R für Reciprocity und C für Competition.

ob er den erwarteten Wert seiner Motivationsfunktion (oder Nutzenfunktion)  $v_i = v_i(y_i, \sigma_i)$  maximiert, wobei  $y_i$  die monetäre Auszahlung von Akteur  $i$ ,

$$\sigma_i = \sigma_i(y_i, c, n) = \begin{cases} y_i / c, & \text{falls } c > 0 \\ 1/n, & \text{falls } c = 0 \end{cases}$$

der relative Anteil von  $i$  an der Gesamtauszahlung,  $c$  = die Gesamtauszahlung aller Akteure (oder der ‚Kuchen‘) und  $n$  die Anzahl aller Akteure ist. Anders als im Standardmodell wird angenommen, dass Akteure nicht nur durch ihre monetäre Auszahlung  $y_i$  („mehr Geld ist besser als weniger Geld“) motiviert sind, sondern auch durch ihre relative Position  $\sigma_i$  in der Gruppe. Dabei wird angenommen, dass die Akteure darunter leiden, wenn sie weniger als die Anderen (im Mittelwert) haben, aber auch wenn sie mehr haben. Der Disnutzen von Ungleichheit kann für die Akteure sehr unterschiedlich sein; während einige keinen Wert auf soziale Vergleiche legen mögen, mag für andere lediglich unvorteilhafte Ungleichheit problematisch sein, während wiederum andere jede Art von Ungleichheit ablehnen.<sup>7</sup> Die Heterogenität individueller sozialer Präferenzen ist eine entscheidende Komponente des Modells, die für die Abbildung bestimmter Makro-Phänomene notwendig ist. Das Modell kann zwar individuelles soziales Verhalten oft nicht prognostizieren (wie viel gibt Lieschen Müller im Diktatorspiel ab?), es sagt aber sehr wohl voraus, wie sich soziales Verhalten in der Gruppe insgesamt ausbildet (siehe Bolton und Ockenfels 2000 für die Details und mathematische Formulierungen).

Es gibt wohl kaum eine einfachere Möglichkeit, soziale Vergleiche im Standardmodell zu berücksichtigen als durch die Modifikation des ERC-Modells. Umso überraschender ist, dass sich diese Motivationsstruktur zusammen mit dem Gleichgewichtskonzept als außerordentlich erfolgreich erwiesen hat, scheinbar disparate Verhaltensweisen auf ein gemeinsames Prinzip zurück führen zu können. Gemäß dem Modell gibt ein Akteur, der soziale Ungleichheit ablehnt, in Diktatorspielen etwas ab, aber nicht mehr als die Hälfte. In Ultimatumspielen werden faire Aufteilungen immer angenommen, aber unfairere Aufteilungen mit höherer Wahrscheinlichkeit abgelehnt; die im Standardmodell prognostizierte Gleichgewichtsaufteilung wird stets abgelehnt. Im Ergebnis geben Proposer im Ultimatumspiel aus strategischen Gründen (nämlich aus Furcht vor Ablehnung) mehr ab als im Diktatorspiel. Auch prognostiziert das Modell konditional kooperatives (reziprokes) Verhalten im Gefangenendilemma und in zahlreichen anderen Dilemmata wie Vertrauensspielen (Berg et al. 1995; Bolton et al. 2004) oder dem gift-exchange Spiel von Fehr et al. (1993). Auf der anderen Seite ist das Modell jedoch auch konsistent mit wettbewerblichem Verhalten im Marktspiel; ein Käufer, der die Konkurrenz meidet und sich nicht auf den Bietkampf einlässt, steht nicht nur absolut gesehen (nämlich finanziell) schlecht da, sondern auch bezüglich seiner relativen Position: er müsste dann nämlich zusehen, wie andere sich den Kuchen aufteilen. Im Ergebnis wird sich niemand dem Wettbewerb entziehen, unabhängig davon, wie sozial die Präferenzen auch sein mögen.

All diese Prognosen sind im Einklang mit den empirischen Befunden, teilweise auch bis ins Detail. Dies legt nahe, dass aus den robusten empirischen Ergebnissen nicht etwa folgt, dass soziale Präferenzen kontextabhängig „aufquellen“. Vielmehr führt das Zusammenspiel eigennütziger und sozialer Präferenzen mit den institutionellen und sozialen Rahmenbedingungen zu systematisch unterschiedlichen

---

<sup>7</sup> Das Modell unterstellt nicht, dass die Ausprägung der eigenen sozialen Präferenzen anderen Akteuren bekannt ist.

und prognostizierbaren Makro-Phänomenen in unterschiedlichen Kontexten. Beispielsweise ist es in Verhandlungen zuweilen selbst für Egoisten klug, sich aus strategischen Gründen fair zu verhalten, und andererseits ist es im Wettbewerb selbst für Akteure mit ausgeprägten sozialen Präferenzen nicht vermeidbar, sich dem Konkurrenzdruck zu beugen (für eine aktuelle ausführliche und kritische Würdigung des Modells, siehe Cooper und Kagel, im Erscheinen, und z.B. Fehr und Schmidt 2006 für eine ähnliche Schlussfolgerung).

Zugleich ermöglicht die mathematisch stringente Modellierung sozialer Präferenzen und die dadurch ermöglichte Ableitung falsifizierbarer Verhaltenshypothesen unter verschiedenen sozialen und ökonomischen Bedingungen auch Aussagen über die zugrundeliegende Natur sozialen Verhaltens, beispielsweise über den inhärenten Zusammenhang zwischen Fairness und Reziprozität in vielen experimentellen Befunden (z.B. Bolton et al. 2008) oder über die Selbstzentriertheit sozialer Verhaltensweisen. Ein Beispiel für letzteres ist das Güth-van Damme (1998) Spiel, ein Ultimatumspiel, in dem Proposer und Responder ihren Kuchen zusätzlich auf einen Dummy aufteilen, der in dem Aufteilungsprozess jedoch nichts zu sagen hat. Wie im Ultimatumspiel macht der Proposer einen Vorschlag und der Responder nimmt an oder lehnt ab.

Abbildung 4 ungefähr hier.

Abbildung 4 zeigt das Ergebnis: der Proposer bietet dem Dummy praktisch nichts und dem Responder seinen fairen Anteil (nämlich ein Drittel) an. Güth und van Damme (1998) schreiben, dass sie in ihren Daten merkwürdigerweise weder Evidenz für Fairness noch für Eigennutz finden. Eigennutz würde implizieren, dass weder der Responder noch der Dummy einen signifikanten Anteil am Kuchen bekommen (siehe „Gleichgewicht“ unter der Eigennutzannahme in der Abbildung), und fair wäre es gewesen, wenn der Proposer auch dem Dummy seinen fairen Anteil gibt („Gleichheit“ in der Abbildung), oder wenn zumindest der Responder unfaire Aufteilungen nicht nur für sich, sondern auch für den Dummy ablehnt. Doch Abbildung 4 zeigt, dass dies nicht passiert.<sup>8</sup> Tatsächlich sind diese Beobachtungen jedoch mit dem ERC-Modell und damit auch mit Fairness in Ultimatumspielen und Eigennutz im Wettbewerb konsistent, wie die mathematische Analyse aufzudecken vermag. Der Grund steckt in der ERC Motivationsfunktion: die Akteure kümmern sich lediglich um ihre *eigene* relative Position  $\sigma_i$ , nicht aber um die relative Position der anderen Akteure. Entsprechend lehnt der Responder niemals ab, wenn zumindest er ‚seinen fairen‘ Anteil erhält, und der Proposer sieht sich folglich nicht genötigt, den Dummy zu bedienen (siehe Bolton und Ockenfels 1998 für eine mathematische und statistische Analyse dieser und weiterer Effekte). Das Ergebnis erscheint zwar normativ unfair, ist aber konsistent mit einem in anderen Kontexten entwickelten deskriptiven Modell rationalen, selbstzentrierten sozialen Verhaltens.

---

<sup>8</sup> Güth und van Damme (1998: 242) schreiben: „The experimental data clearly refute the idea that proposers are intrinsically motivated by considerations of fairness: They only allocate marginal amounts to the dummy“. Und (1998: 230): „there is not a single rejection that can be clearly attributed to a low share for the dummy.“

## 5. Zusammenfassung und Ausblick

In diesem Beitrag haben wir zunächst skizziert, dass das Standardmodell des Rational Choice-Ansatzes, das zusätzlich zur Rationalitätsannahme auch die Eigennutzannahme verwendet, mit manchen experimentellen empirischen Befunden unvereinbar ist. Zwar erklärt das Standardmodell viele empirische Regelmäßigkeiten, die wir in experimentellen Märkten antreffen, zahlreiche empirische Regelmäßigkeiten in Experimenten zu Ultimatum- und Diktatorspielen und zum Gefangenendilemma widersprechen aber dem Standardmodell. Wir haben sodann Alternativen zum Standardmodell präsentiert, in denen auf die Eigennutzannahme verzichtet wird. Unsere Alternativen lösen das Problem, dass andere Motivationsannahmen als die Eigennutzannahme leicht zur „Immunisierung“ (Albert 1977) von Rational Choice-Modellen führen können. Mit einem Modell endogener Präferenzänderungen haben wir Bedingungen spezifiziert, unter denen rationale und eigeninteressierte Akteure ihre eigenen Präferenzen modifizieren werden. Andere als eigennützige Präferenzen, genauer: Assurance Game Präferenzen, werden in diesem Modell nicht *angenommen*, sondern sind *Konsequenzen* des Modells. Das ERC-Modell ist dagegen ein Modell, das mit einer relativ „sparsamen“ Alternative zur Eigennutzannahme, nämlich der Annahme, dass Akteure sowohl durch ihre eigenen materiellen Auszahlungen als auch durch ihre relative Position motiviert sind, zahlreiche empirische Regelmäßigkeiten aus experimentellen Spielen organisieren kann. Das ERC-Modell kann wie das Standardmodell eigennütziges Verhalten in Marktspielen erklären, aber es ist dem Standardmodell überlegen, weil es auch faires Verhalten in Ultimatum- und Diktatorspielen sowie reziprokes Verhalten im Gefangenendilemma erklären kann, und zwar gerade auch bei anonymen und nicht wiederholten Interaktionen.

Man kann sich im Übrigen klar machen, dass es systematische Zusammenhänge zwischen den Assurance Game Präferenzen gibt, die von rationalen Akteuren in unserem Modell endogener Präferenzänderungen gewählt werden, und ERC Motivationsfunktionen, die im Gefangenendilemma zu wechselseitiger Kooperation führen. Ein rationaler Akteur mit einer ERC Motivationsfunktion kann sich nämlich im Kontext des Gefangenendilemmas aus Kapitel 3 lediglich entweder wie durch eigennützigen Präferenzen induziert oder wie durch Assurance Game Präferenzen induziert verhalten (siehe Bolton und Ockenfels 2000, Kapitel VI.A für Details). In diesem Sinn kann man also ERC Motivationsannahmen im Gefangenendilemma-Kontext auch als Implikationen des Modells endogener Präferenzänderungen auffassen (ähnliche Zusammenhänge gibt es auch zwischen Assurance Game Präferenzen und Nutzenfunktionen à la Fehr und Schmidt 1999, vgl. z.B. Diekmann und Voss 2008).

Zugleich ist klar, dass solche Modelle sozialer Präferenzen nicht alle Verhaltensphänomene richtig erfassen können. Soziales Verhalten ist ein komplexes Phänomen, das durch motivationale Prinzipien allein nicht vollständig erfasst werden kann, sondern auch kognitive, biologische (einschließlich neurologische, chemische und quantenphysikalische), adaptive und andere Wurzeln besitzt; z.B. kann soziales Verhalten auch durch das Einsprühen bestimmter Stoffe in die Nase oder durch das kurzzeitige Halten einer Tasse Kaffee beeinflusst werden (Kosfeld et al. 2005, Williams und Bargh 2008).<sup>9</sup> Doch es ist oft nicht hilfreich, und in der Regel auch

---

<sup>9</sup> Die skizzierten Modelle können naturgemäß auch nicht alle potenziell relevanten Motivationen abbilden; siehe Cooper und Kagel (im Erscheinen), Bolton und Ockenfels (2006, 2009, im Erscheinen b) sowie die dort zitierten Referenzen für wichtige Beispiele und Diskussionen (etwa zur Rolle

nicht möglich, alle potenziell relevanten Faktoren gleichzeitig zu berücksichtigen. Modelle, seien sie mathematisch oder verbal formuliert, müssen generell von der Realität abstrahieren, um nützlich zu sein. Das richtige Abstraktionsniveau hängt von der Fragestellung ab (siehe Lindenberg 1992, Bolton und Ockenfels 2009, Bolton, im Erscheinen, und Roth 1996 für ähnliche Argumente).

Es hilft, wenn man sich Theorien sozialen Verhaltens als Landkarten vorstellt. Um die Bewegung von Planeten zu verstehen, reicht es in der Regel aus, sich die ganze Erde als einen Massepunkt vorzustellen – die einfachste Version einer Landkarte. Einem Bergsteiger ist mit so einer Karte jedoch wenig geholfen. Ebenso kann ein U-Bahn Plan helfen, die richtige U-Bahn zu finden, den Autofahrer dagegen oft in die falsche Richtung schicken. Letzterer benötigt eine detailliertere Karte, doch eine Karte, die die Erdoberfläche bis zum letzten Grashalm abbildet, ist wenig hilfreich, wenn man von Köln nach Utrecht fahren möchte.

Ähnliches gilt für Theorien sozialen Verhaltens. Keine uns bekannte Theorie fängt gleichzeitig alle potenziell relevanten ökonomischen, psychologischen und biologischen Einflussfaktoren sozialen Verhaltens ein; alle Theorien werden demnach streng genommen falsifiziert (soweit sie falsifizierbar sind). Das heißt jedoch nicht, dass die Modelle unnütz sind. Die Erde ist auch kein Massepunkt, und der U-Bahn Plan versagt, wenn er den Fußweg zu unseren Büros weisen soll. Doch für viele Fragen, mit denen sich Soziologen und Ökonomen beschäftigen, ist es allenfalls von untergeordneter Relevanz, ob ein warmes Getränk im Durchschnitt zu mehr Vertrauen führt als ein kaltes Getränk. Eine Theorie, die die Komplexität der Realität reduziert, kann durchaus nützlich sein, da sie Wissenschaftlern und Praktikern Orientierung gibt, so wie die Landkarte Orientierung gibt. Wir glauben, dass nicht-eigennützige Motivationen in vielen Fällen für das Verständnis sozialen Verhaltens relevant sind, was erklärt, warum sie auf vielen Landkarten verzeichnet sind. Diese Landkarten können mit Karten, die andere ökonomische, soziologische, psychologische und biologische Erkenntnisse zusammen stellen, übereinandergelegt werden und im wissenschaftlichen Wettbewerb revidiert, verbessert und zusammengeführt werden.

In empirischer Hinsicht haben wir uns in diesem Beitrag auf experimentelle Evidenzen konzentriert, und zwar auf Evidenzen aus Laborexperimenten. Dafür gibt es gute Gründe (s.o.), aber es ist klar, dass Laborexperimente von vielen Kontexteigenschaften abstrahieren. Gerade dadurch können Hypothesen systematisch überprüft und relevante Einflussfaktoren identifiziert und analysiert werden. Gerade dadurch sind aber auch komplementäre empirische Designs – z.B. Feldexperimente, Vignetten-Studien, Surveystudien, die Analyse nicht-reaktiver Verhaltensdaten und Fallstudien – nützlich, um zu robusten Befunden über Fairness, Reziprozität und Eigennutz in unterschiedlichen Kontexten wie Märkten, sozialen Situationen mit Verteilungsproblemen und sozialen Dilemmata zu kommen (vgl. auch Diekmann 2008; Greiner und Ockenfels 2009; Harrison und List 2004; Levitt und List 2007).

---

prozeduraler Fairness, die das ERC-Modell nicht einfangen kann; siehe dazu aus soziologischer Sicht z.B. Vieth 2009).



## Literatur

- Albert, Hans. 1977. Individuelles Handeln und soziale Steuerung. Die ökonomische Tradition und ihr Erkenntnisprogramm. In *Handlungstheorien – interdisziplinär IV: Sozialwissenschaftliche Handlungstheorien und spezielle systemwissenschaftliche Ansätze*, Hrsg. Hans Lenk, 177-225. München: Fink.
- Becker, Gary S. 1976. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.
- Berg, J., John Dickhaut und Kevin McCabe. 1995. Trust, Reciprocity and Social History. *Games and Economic Behavior* 10: 122–142.
- Bolton, Gary E. Forthcoming: Testing models and internalizing context: A comment on Vernon Smith's 'Theory and experiment: What are the questions?' *Journal of Economic Behavior and Organization*.
- Bolton, Gary E., und Axel Ockenfels. 1998. Strategy and Equity: An ERC-Analysis of the Güth-van Damme Game. *Journal of Mathematical Psychology* 42(2): 215-226.
- Bolton, Gary E., und Axel Ockenfels. 2000. ERC: A Theory of Equity, Reciprocity and Competition *American Economic Review* 90: 166-193.
- Bolton, Gary E., und Axel Ockenfels. 2006. Measuring Efficiency and Equity Motives - A Comment on "Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments". *American Economic Review* 96: 1906-1911.
- Bolton, Gary E., und Axel Ockenfels. 2009. *Testing and Modeling Fairness Motives*. Working Paper.
- Bolton, Gary E., und Axel Ockenfels. Forthcoming: The Limits of Trust. In *Trust and Reputation*, Hrsg. Chris Snijders. New York: Russell Sage.
- Bolton, Gary E., und Axel Ockenfels. Forthcoming: Risk Taking and Social Comparison – A Comment on „Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States". *American Economic Review*.
- Bolton, Gary E., Jordi Brandts und Axel Ockenfels. 2005. Fair Procedures: Evidence from Games Involving Lotteries. *Economic Journal* 115(506): 1054–1076.
- Bolton, Gary E., Jordi Brandts, Elena Katok, Axel Ockenfels und Rami Zwick. 2008. Testing Theories of Other-regarding Behavior - A Sequence of Four Laboratory Studies. In *Handbook of Experimental Economics Results Vol. I.*, Hrsg. Charles R. Plott, Vernon L. Smith, 488-499. North Holland: Elsevier.
- Bolton, Gary E., Elena Katok und Axel Ockenfels. 2004. Trust among Internet Traders – A Behavioral Economics Approach. *Analyse und Kritik* 26: 185-202.
- Brennan, Geoffrey, und James M. Buchanan. 1985. *The Reason of Rules. Constitutional Political Economy*. Cambridge: Cambridge University Press.
- Buskens, Vincent, und Werner Raub. 2010. Rational Choice Research on Social Dilemmas. Erscheint in *Handbook of Rational Choice Social Research*, Hrsg. Rafael Wittek, Tom A. B. Snijders und Victor Nee. New York: Russell Sage.
- Camerer, Colin F. 2003. *Behavioral Game Theory. Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
- Coleman, James S. 1987. Psychological Structure and Social Structure in Economic Models. In *Rational Choice. The Contrast between Economics and Psychology*, Hrsg. Robin M. Hogarth, Melvin W. Reder, 181-185. Chicago: University of Chicago Press.
- Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University Press.

- Cooper, David J., und John H. Kagel. Forthcoming. Other Regarding Preferences: A Selective Survey of Experimental Results. In *Handbook of Experimental Economics, Volume 2*, Hrsg. John H. Kagel, Alvin E. Roth. Princeton, NJ: Princeton University Press.
- Dawes, Robyn M., und Richard H. Thaler. 1988. Anomalies: Cooperation. *Journal of Economic Perspectives* 2: 187-197.
- Diekmann, Andreas. 2008. Soziologie und Ökonomie: Der Beitrag experimenteller Wirtschaftsforschung zur Sozialtheorie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 60: 528-550.
- Diekmann, Andreas. 2009. *Spieltheorie*. Einführung, Beispiele, Experimente. Reinbek bei Hamburg: Rowohlt.
- Diekmann, Andreas, und Thomas Voss (Hrsg.). 2004. *Rational-Choice Theorie in den Sozialwissenschaften. Anwendungen und Probleme*. München: Oldenbourg.
- Diekmann, Andreas, und Thomas Voss. 2008. Soziale Normen und Reziprozität. In *Rational Choice: Theoretische Analysen und empirische Resultate*, Hrsg. Andreas Diekmann, Klaus Eichner, Peter Schmidt, Thomas Voss, 83-100. Wiesbaden: VS Verlag.
- Esser, Hartmut. 1993. *Soziologie. Allgemeine Grundlagen*. Frankfurt a. M.: Campus.
- Esser, Hartmut. 2010. *Das Modell der Frame-Selektion: Eine allgemeine Handlungstheorie für die Sozialwissenschaften?*, in diesem Band.
- Fehr, Ernst, und Herbert Gintis. 2007. Human Motivation and Social Cooperation: Experimental and Analytical Foundations. *Annual Review of Sociology* 33: 43-64.
- Fehr, Ernst, Georg Kirchsteiger und Arno Riedl. 1993. Does Fairness Prevent Market Clearing: An Experimental Investigation. *Quarterly Journal of Economics* 108: 437-459.
- Fehr, Ernst, und Klaus Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114, 817-868.
- Fehr, Ernst, und Klaus M. Schmidt. 2006. The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories. In *Handbook of the Economics of Giving, Altruism and Reciprocity, Volume I*, Hrsg. Serge-Christophe Kolm, Jean Mercier Ythier, 615-691. Amsterdam: Elsevier.
- Forsythe, Robert, Joel L. Horowitz, N.E. Savin und Martin Sefton.. 1994. Fairness in Simple Bargaining Experiments. *Games and Economic Behavior* 6: 347-369.
- Friedman, Milton. 1953. *Essays in Positive Economics*. Chicago, IL: University of Chicago Press.
- Goldthorpe, John H.. 1996. The Quantitative Analysis of Large-scale Data Sets and Rational Action Theory: For a Sociological Alliance. *European Sociological Review* 12: 109-126.
- Greiner, Ben, und Axel Ockenfels. 2009. Vom Labor ins Feld: Die Ökonomik des Vertrauens. Erscheint in *Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 49: Wirtschaftssoziologie*.
- Güth, Werner, und Eric van Damme. 1998. Information, Strategic Behavior and Fairness in Ultimatum Bargaining: An Experimental Study. *Journal of Mathematical Psychology* 42: 227-247.
- Güth, Werner, und Hartmut Kliemt. 2000. Evolutionarily Stable Co-operative Commitments. *Theory and Decision* 49: 197-221.
- Güth, Werner, und Axel Ockenfels. 2003. The Coevolution of Trust and Institutions in Anonymous and Non-anonymous Communities. In *Jahrbuch für Neue*

- Politische Ökonomie* 20. Hrsg. Manfred J. Holler, Hartmut Kliemt, Dieter Schmidtchen, Manfred Streit, 157-174. Tübingen: Mohr.
- Güth, Werner, und Axel Ockenfels. 2005. The Coevolution of Morality and Legal Institutions: An Indirect Evolutionary Approach. *Journal of Institutional Economics* 1(2): 155-174.
- Güth, Werner, Rolf Schmittberger und Bernd Schwarze, 1982. An Experimental Analysis of Ultimatum Bargaining, *Journal of Economic Behavior and Organization*, 3, 367-388.
- Harrison, Glenn W., und John A. List. 2004. Field Experiments. *Journal of Economic Literature* 42: 1009-1055.
- Harsanyi, John C. 1976. *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht: Reidel.
- Harsanyi, John C. 1977. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.
- Hedström, Peter. 2005. *Dissecting the Social. On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
- Kosfeld, Michael, Markus Heinrichs, Paul J. Zak, Urs Fischbacher und Ernst Fehr. 2005. Oxytocin Increases Trust in Humans. *Nature* 435: 673-676.
- Levitt, Steven D., und John A. List. 2007. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?. *Journal of Economic Perspectives* 21(2): 153-174
- Lindenberg, Siegwart. 1992. The Method of Decreasing Abstraction. In *Rational Choice Theory. Advocacy and Critique*, Hrsg. James S. Coleman, Thomas J. Fararo, 3-20. London: Sage.
- Lindenberg, Siegwart, und Reinhard Wippler. 1978. Theorienvergleich: Elemente der Rekonstruktion. In *Theorienvergleich in den Sozialwissenschaften*, Hrsg. Karl-Otto Hondrich, Joachim Matthes, 219-231. Darmstadt: Luchterhand.
- Mas-Colell, Andreu, Michael D. Whinston und Jerry R. Green. 1995. *Microeconomic Theory*. New York: Oxford University Press.
- Ockenfels, Axel. 1999. *Fairness, Reziprozität und Eigennutz*. Tübingen: Mohr.
- Ockenfels, Axel. 2009. Marktdesign und Experimentelle Wirtschaftsforschung. *Perspektiven der Wirtschaftspolitik* 10 (Special Issue): 31-54.
- Opp, Karl-Dieter. 1979. *Individualistische Sozialwissenschaft*. Stuttgart: Enke.
- Popper, Karl R.. 1973[1934]. *Logik der Forschung*. 5. Aufl.. Tübingen: Mohr.
- Popper, Karl R.. 1974[1963]. *Conjectures and Refutations. The Growth of Scientific Knowledge*. 5. Aufl.. London: Routledge and Kegan Paul.
- Raub, Werner. 1990. A General Game-Theoretic Model of Preference Adaptations in Problematic Social Situations. *Rationality and Society* 2: 67-93.
- Raub, Werner, und Vincent Buskens. 2006. Spieltheoretische Modelle und empirische Anwendungen in der Soziologie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 44: Methoden der Sozialforschung*, 560-598.
- Raub, Werner, und Thomas Voss. 1990. Individual Interests and Moral Institutions: An Endogenous Approach to the Modification of Preferences. *Social Institutions: Their Emergence, Maintenance, and Effects*, Hrsg. Michael Hechter, Karl-Dieter Opp, Reinhard Wippler, 81-117. New York: Aldine.
- Roth, Alvin E. 1988. Laboratory Experimentation in Economics: A Methodological Overview. *Economic Journal* 98: 974-1031.
- Roth, Alvin E. 1995. Bargaining Experiments. In *Handbook of Experimental Economics*, Hrsg. John H. Kagel, Alvin E. Roth, 253-348. Princeton: Princeton University Press.

- Roth, Alvin E. 1996. Comments on Tversky's 'Rational Theory and Constructive Choice'. In *The Rational Foundations of Economic Behavior*, Hrsg. Kenneth Arrow, Enrico Colombatto, Michael Perlman, Christian Schmidt, 198-202. London-New York: Macmillan.
- Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, und Shmuel Zamir. 1991. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Review* 81, 1068-1095.
- Schmidt, Klaus. 2009. The Role of Experiments for the Development of Economic Theories. *Perspektiven der Wirtschaftspolitik* 10 (Special Issue), 14-30.
- Schüßler, Rudolf. 1988. Der homo oeconomicus als skeptische Fiktion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 40, 447-463.
- Sen, Amartya. 1982. *Choice, Welfare and Measurement*. Oxford: Blackwell.
- Vieth, Manuela. 2009. *Commitments and Reciprocity. Experimental Studies on Obligation, Indignation, and Self-Consistency*. PhD thesis ICS-Utrecht.
- Weber, Max. 1976 (1921). *Wirtschaft und Gesellschaft*. 5. Aufl.. Tübingen: Mohr.
- Williams, Lawrence E., und John A. Bargh. 2008. Experiencing Physical Warmth Promotes Interpersonal Warmth. *Science* 322 (5901): 606-607.

		Akteur 2	
		Kooperation	Defektion
Akteur 1	Kooperation	$R,R$	$S,T$
	Defektion	$T,S$	$P,P$

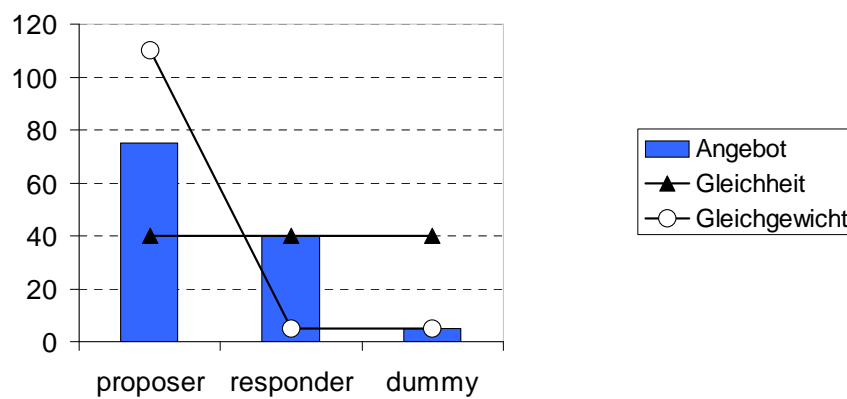
**Abbildung 1: Das Gefangenendilemma** ( $T > R > P > S$ )

		Akteur 2	
		Kooperation	Defektion
Akteur 1	Kooperation	$R_1^e, R_2^e$	$S_1^e, T_2^e$
	Defektion	$T_1^e, S_2^e$	$P_1^e, P_2^e$

**Abbildung 2: Das effektive Spiel**

		Akteur 2	
		$O_2^{AG}$	$O_1^D$
Akteur 1	$O_1^{AG}$	$R^n, R^n$	$P^n, P^n$
	$O_1^D$	$P^n, P^n$	$P^n, P^n$

**Abbildung 3: Das Spiel M** ( $R^n > P^n$ )



**Abbildung 4: Ergebnisse im Güth-van Damme Spiel**